

# **Eyal Orbach - id 015369317**

# **Daniel Juravski - id 206082323**

- **Can the two languages be distinguished using a bag-of-words approach? Explain why.**

Bag of words can accept an input of inconsistent length but does not maintain its order, therefore would fail to recognize the convention.

- **Can the two languages be distinguished using a bigram or trigram based approach? Explain why.**

For any n-gram we choose, the sequence of digits can be larger than that chosen n, having it larger than our n-gram window will prevent our network from recognizing the rules here, because the order of the letters would be out of the window's boundaries and without memory there would be no way for the model to distinguish between a positive and a negative example.

- **Can the two languages be distinguished using a convolutional neural network? Explain why.**

CNN requires fixed length for the input, or at least a maximal length where anything smaller will be padded. Since the strings are of variant length any window size we choose as the fixed input size, can be too small to capture the phenomena without memory.